JOURNAL OF APPROXIMATION THEORY 21, 60–88 (1977)

# On the Deterministic and Stochastic Approximation of Regions*

P. J. DAVIS, R. A. VITALE,[1] AND E. BEN-SABAR

*Brown University, Providence, Rhode Island 02912*

*Communicated by O. Shisha*

Received January 2, 1976

## 1. INTRODUCTION

The problem of approximating a given two-dimensional region by a simpler region, selected from a family of simpler regions, occurs with reasonable frequency. One might point to Kepler's problem of approximating the observed positions of a planet by an ellipse. It is often possible by a variety of devices to reduce the problem to a one-dimensional one and apply well-established methods of one-dimensional approximation theory.

In this paper, however, we shall stress methods that are intrinsically two-dimensional. As a concrete example, we shall concentrate on the problem of approximating a given triangle by a circle. This problem is already sufficiently complicated so that closed form expressions may be very difficult to obtain, but sufficiently simple so that the main features are not obscured. In dealing with general situations, we shall have no great interest in "pathological" regions, but shall assume that all our regions have piecewise analytic (or even simpler) boundaries. Very often the family of approximants consists of one basic figure $R$ together with all its transforms $gR$ where $g$ is an element of some familiar group $G$ of plane transformations. Thus, for examples, the approximants might consist of (a) all circles in the plane, (b) all ellipses, (c) all figures congruent to a given figure $R$, etc.

## 2. THE METHOD OF INTERPOLATION

Let $R_1$ and $R_2$ be two plane regions. By analogy with the one-dimensional case, we shall say $R_1$ *interpolates* to $R_2$ (or vice versa) at the points $P_1$, $P_2$, ...,

* Results shown in this paper were obtained, in part, in the course of research sponsored by the National Science Foundation under Grants MPS-75-07468 and GP-40521.

[1] Current address: Mathematics Research Center, University of Wisconsin, Madison, Wis. 53706.

$P_n$ if these points lie in both $R_1$ and $R_2$ : $P_i \in R_1 \cap R_2$, $i = 1, 2,..., n$. While this definition is by no means devoid of interest, we shall in actuality deal with the case where each of the points $P_1 ,..., P_n$ lies on the boundaries of both $R_1$ and $R_2$.

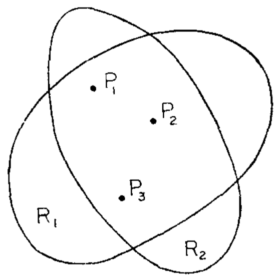


FIGURE 1.

One may also consider interpolatory conditions of *osculatory Hermite* type. If at a point $P_1$ common to the boundaries $\partial R_1$, $\partial R_2$ of $R_1$, $R_2$, both $\partial R_1$ and $\partial R_2$ have a direction, and if these directions coincide, then we shall write $(\partial R_1)' = (\partial R_2)'$ at $P = P_1$, and consider that *two* interpolatory conditions are fulfilled at $P_1$.

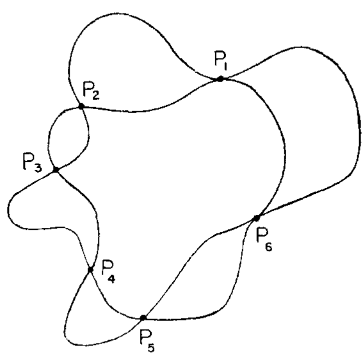


FIGURE 2.

We turn immediately to the case $R_1 = T =$ triangle, $R_2 = C =$ circle. A circle is determined uniquely by three noncollinear points. Therefore, given three points $P_1$, $P_2$, $P_3$ on $\partial T$ not all lying on one side of $T$, there exists a unique circle $C$ which interpolates to $T$ at the $P_i$.

Some notable selections are

(a) select $P_i$ as the vertices of $T$. This leads to the *circumscribed* circle;

(b) select $P_i$ as the feet of the perpendiculars drawn from the incenter of $T$. This leads to the *inscribed* circle;

(c) take $P_i$ as the midpoints of the sides of $T$. This circle also passes through the feet of the altitudes of $T$ and through three other distinguished points. It is called *the nine-point-circle* of $T$.

Note that the circle in (b) has second-order contact with $\partial T$ at $P_i$ and so satisfies six interpolatory conditions.

(d) In addition to the nine-point circle, there are numerous other interpolatory circles which occur in advanced synthetic geometry. (See, e.g., Johnson [11].) For example, given a (nondegenerate) triangle $T$, there is a unique point $K$ which minimizes the sum of the squares of the distances from the point to the sides of $T$ (see, e.g., [11, p. 213]). $K$ is known as the *Lemoine point* or the *symmedian point* of $T$. Through $K$ draw lines antiparallel to the three sides of $T$. These lines intersect the sides of $T$ in six points which lie on a common circle known as the *second Lemoine circle.* We shall have occasion to mention the symmedian point and elaborate the definitions later. (See Example 1, Section 6).

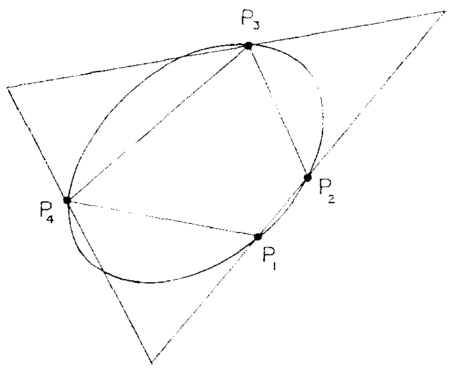


FIGURE 3.

The case $R_1 = T =$ triangle, $R_2 = E =$ ellipse, is also of interest. Here the situation is different in that we have to dispense with uniqueness. The equation of a conic section, $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$, contains five essential parameters. Hence five points in the plane of which no three are collinear will determine a unique conic (not necessarily an ellipse). Three points on $\partial T$, not all on one side of $T$ determine a circle, hence an ellipse. It can be shown that given four points, $P_1$, $P_2$, $P_3$, $P_4$, which are the vertices of a nondegenerate convex quadrilateral, a one parameter family of ellipses may be found passing through these points; hence, given four points on a triangle, two on side 1, one on side 2, one on side 3, the same may be asserted.

A method of interpolation which has become very popular, particularly in computer-aided design, is the use of *parametric, periodic, interpolatory cubic* *splines*. Given $n$ points in the plane $P_1, P_2, \ldots, P_n, P_{n+1} = P_1$, and $n+1$ parametric values $0 = t_1 < t_2 < \cdots < t_{n+1}$, we can find two functions $x = x(t)$, $y = y(t)$, each of which is a cubic spline of period $t_{n+1}$ and such that $P_i = (x(t_i), y(t_i))$, $i = 1, 2, \ldots, n$. As this method is strongly one-dimensional, we mention it, but shall not pursue it further.

## 3. The Method of Moments

A second interpolatory process occurs when moments are used instead of functional values. By the moments of a bounded region $R$ are meant the numbers

$$\mu_{m,n}(R) = \mu_{m,n} = \iint_R x^m y^n \, dx\, dy, \qquad m, n = 0, 1, \ldots. \tag{3.1}$$

The complex form of these moments has also been found convenient

$$\tau_{m,n}(R) = \tau_{m,n} = \iint_R z^m \bar{z}^n \, dx\, dy, \qquad z = x + iy, \bar{z} = x - iy. \tag{3.1a}$$

More generally, one might work with

$$\mu_{m,n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^m y^n C_R(x, y)\, w(x, y)\, dx\, dy, \qquad m, n = 0, 1, \ldots, \tag{3.2}$$

where $C_R(x, y)$ is the characteristic function of the set $R$ ($C_R(x, y) = 1$ if $(x, y) \in R$, $C_R(x, y) = 0$ otherwise) and where $w(x, y) \geqslant 0$ is an appropriate fixed weighting function. The discrete Fourier transform (or trigonometric moments)

$$\mu_{m,n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_R(x, y)\, e^{2\pi i m x} e^{2\pi i n y}\, dx\, dy \tag{3.3}$$

might also be considered. Other possibilities are the discrete Walsh transform, etc.

As particular instances of (3.1), we cite

$$\mu_{0,0} = \iint_R dx\, dy = \text{area of } R = A(R) = A,$$

$$\mu_{1,0} = \iint_R x\, dx\, dy = y\text{-moment of } R;\ x_{\text{c.g.}} = (1/A)\, \mu_{1,0},$$

$$\mu_{0,1} = \iint_R y\, dx\, dy = x\text{-moment of } R,\ y_{\text{c.g.}} = (1/A)\, \mu_{0,1}$$

($x_{\text{c.g.}}$ and $y_{\text{c.g.}}$ are the coordinates of the center of gravity of $R$), (3.4)

$$\mu_{2,0} = \iint_R x^2\,dx\,dy = \text{moment of interia of } R \text{ about the } y\text{-axis},$$

$$\mu_{0,2} = \iint_R y^2\,dx\,dy = \text{moment of inertia of } R \text{ about the } x\text{-axis},$$

$$\mu_{1,1} = \iint_R xy\,dx\,dy = \text{product of inertia of } R.$$

The method of moments proceeds by *moment matching*. That is, one approximates a region $R$ from among a family of regions $\mathscr{S}$ by selecting a region $S \in \mathscr{S}$ for which

$$\mu_{i,j}(R) = \mu_{i,j}(S) \qquad \text{for some finite number of index pairs } (i,j). \tag{3.5}$$

EXAMPLE. For a given triangle $T$ there is one and only one circle $C$ such that area $(C)$ = area $(T)$ and c.g. $(C)$ = c.g. $(T)$.

## 4. A GENERALIZATION: FEATURES AND FEATURE MATCHING

Let $\mathscr{R}$ be a family of regions $R$. Suppose that for each $R \in \mathscr{R}$ there is a mapping $\mu$ of $\mathscr{R}$ into the set of real or complex numbers. This mapping will be called a functional on $\mathscr{R}$ or a *feature*. Among the various features which have been found to be of interest are *area $A$, perimeter $L$, the higher trigonometric moments* or *discrete Fourier transform, diameter, connectivity, various measures of symmetry, and aspect ratio.* One should also mention the elementary "*point*" *feature* $\mu_P : \mu_P(R) = 1$ if $P \in R$, $\mu_P(R) = 0$, otherwise. One might also consider very "advanced" features such as *isoperimetric ratio* $(L/A)$, *capacity, torsional rigidity, principal frequency*, etc., etc. (See Pólya and Szegö [16].)

The approximation of a region $R$ by a member $S$ of a family $\mathscr{S}$ of regions may proceed by *feature matching*. That is, let $\mu_i$, $i = 1, 2,..., n$, be a finite number of features. One now requires that an $S \in \mathscr{S}$ be selected so that

$$\mu_i(S) = \mu_i(R) \qquad i = 1, 2,..., n. \tag{4.1}$$

Of course, it may be impossible to meet conditions (4.1), in which case one can proceed by feature matching in some *approximate sense*. We shall elaborate this notion subsequently.

Suppose that $\{\mu_i\}$ designates a finite or a denumerable set of features defined on $\mathscr{R}$, a family of regions. Suppose, further, that for $R, S \in \mathscr{R}$,

$$\mu_i(R) = \mu_i(S) \qquad \text{for all } i \tag{4.2}$$

implies that $R = S$. Then $\{\mu_i\}$ will be said to be *complete in* $\mathscr{R}$. That is to say, the regions of $\mathscr{R}$ can be completely identified by the features $\{\mu_i\}$. Generally speaking, strict equality: $R = S$ is of the most interest for us. But the possibility of weaker equivalences such as $R = S$ a.e. must be allowed for.

EXAMPLE 1. Let $\mathscr{R}$ consist of all nondegenerate triangles in the plane. The four complex moments $\tau_{0,0}$, $\tau_{1,0}$, $\tau_{2,0}$, $\tau_{3,0}$ form a complete set of features in $\mathscr{R}$. See Davis [5].

EXAMPLE 2. Let $\mathscr{R}$ consist of all nondegenerate triangles together with all nondegenerate circles. Then the same conclusion holds.

The classic cases of the completeness (or closure) of a system of functions are examples of the completeness of a system of "features." Among the best known complete sets of functions is the set of powers and correspondingly, we have the following theorem.

*Let B and D be open bounded sets in the plane which possess exterior points in the neighborhood of any boundary point. Then,*

$$\iint_B x^m y^n \, dx \, dy = \iint_D x^m y^n \, dx \, dy, \qquad m, n = 0, 1, \ldots, \tag{4.3}$$

*implies* $B = D$.

(See Davis and Pollak [4].)

Note the conclusion $B = D$; not $B = D$ a.e., this is because we have restricted the nature of the sets $B$ and $D$.

## 5. METRIC SPACES OF REGIONS

Let $\mathscr{R}$ designate a family of regions. It is possible to introduce the notion of distance so that $\mathscr{R}$ becomes a metric space. We first recall the relevant definitions.

Let $d$ be a function from $\mathscr{R} \times \mathscr{R}$ to the set of nonnegative real numbers with the properties

(a) $d(R, R) = 0$ for all $R \in \mathscr{R}$,

(b) $d(R, S) = d(S, R)$ for all $R, S \in \mathscr{R}$,

(c) $d(R, T) \leqslant d(R, S) + d(S, T)$ for all $R, S, T \in \mathscr{R}$.

Such a function is called a *pseudometric* on $\mathscr{R}$ and the pair $(\mathscr{R}, d)$ is called a *pseudometric space*. If, in addition, the function $d(R, S)$ satisfies

(d) $d(R, S) = 0$ implies $R = S$,

then $(\mathscr{R}, d)$ is called a *metric space*.

In some instances it may be sufficient to deal with pseudometric spaces.

There are numerous ways of defining a pseudometric (or a metric) for a family $\mathscr{R}$ of regions.

(a) We begin with an elementary "distance" between two sets. Let $A$ and $B$ be two subsets of the plane, and let $d$ designate any pseudometric in the plane. Then the number

$$D(A, B) = \inf\{d(p, q) : p \in A, q \in B\} \tag{5.1}$$

is commonly referred to as the *distance between the subsets $A$ and $B$*.

As it stands the set of subsets with this distance does not form a pseudometric space. But one can be formed by identifying sets of distance zero. For details see, e.g., Kelley [14, p. 123].

(b) *The Hausdorff metric.* Let $R$ and $S$ be sets of the plane. Let $d$ be any metric in the plane. Let

$$\begin{aligned} H(R, S) &= \max\{\sup_{p \in R} d(p, S), \sup_{p \in S} d(p, R)\} \\ &= \inf\{\epsilon > 0 \mid R \subset S + \epsilon B, S \subset R + \epsilon B\}, \end{aligned}$$

$$B = \text{unit disc with respect to the } d \text{ metric.}$$

Then $H(R, S)$ defines a metric space. For the notation $S + \epsilon B$, see Section 7.

(c) *Metric induced by function spaces.* Let $\mathscr{R}$ designate the family of bounded regions and their closure in the plane.

For each $R \in \mathscr{R}$, designate by $C_R(x, y)$ the characteristic function of the region $R$ ($C_R = 1$ if $(x, y) \in R$, $C_R = 0$ otherwise). Let $C$ be a family of functions defined over the whole plane which contains all the $C_R$, $R \in \mathscr{R}$, and which has been provided with a (pseudo) metric $d(f, g)$. Then, the definition

$$d(R, S) = d(C_R, C_S) \tag{5.3}$$

induces a (pseudo) metric in $\mathscr{R}$.

EXAMPLE 1. Let $C$ be the $L^p$ space on $-\infty < x, y < \infty$, $p \geq 1$, so that

$$d(R, S) = \left( \iint_{-\infty}^{\infty} |C_R(x, y) - C_S(x, y)|^p \, dx \, dy \right)^{1/p}.$$

Let $\psi(x, y) = |C_R(x, y) - C_S(x, y)|^p$. Then, we have

$$\begin{aligned} \psi(x, y) &= 0 \quad \text{if} \quad (x, y) \in \overline{R \cup S}, \\ \psi(x, y) &= 0 \quad \text{if} \quad (x, y) \in R \cap S, \\ \psi(x, y) &= 1 \quad \text{if} \quad (x, y) \in (R - S) \cup (S - R). \end{aligned}$$

Therefore,

$$d^p(R, S) = \text{area}\ (R - S) + \text{area}\ (S - R) = \text{area of symmetric difference of } R \text{ and } S. \tag{5.5}$$

The selection $p = 1$ leads to the particularly simple definition

$$d(R, S) = \text{area}(R - S) + \text{area}(S - R). \tag{5.6}$$

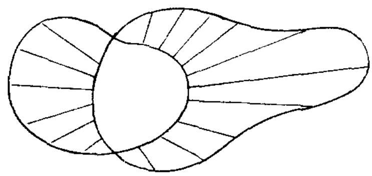

FIGURE 4.

In every case, this leads to a pseudometric space. If a metric space is desired, the family $\mathscr{R}$ should be restricted in such a way that the condition

$$\text{area}(R - S) + \text{area}(S - R) = 0 \tag{5.7}$$

implies $R = S$.

COUNTEREXAMPLE. Let $R =$ the unit disc $x^2 + y^2 < 1$. Let $S =$ the punctured disc $0 < x^2 + y^2 < 1$. Then $\text{area}(R - S) + \text{area}(S - R) = 0$, but $R \neq S$.

One way in which this can be done is to require that the sets of $\mathscr{R}$ are all bounded, open sets which have an exterior point in the neighborhood of any boundary point.

(d) *Metrics induced by features.* Let $\mathscr{R}$ be a set of regions in the plane. Let $\{\mu_i\}$ designate a set of features. For each $R \in \mathscr{R}$, set up the vector (of real or complex numbers) $v(R) = (\mu_1(R), \mu_2(R), \ldots)$. Depending on whether the set of features is finite or denumerably infinite, the vector $v(R)$ will have a finite or an infinite number of components. Assume that the $v(R)$ can be embedded in a normed linear space of appropriate dimension and with norm $\|\ \ \|$. Introduce the pseudometric $d$ by means of

$$d(R, S) = \| v(R) - v(S) \|; \tag{5.8}$$

it is easily shown that $(\mathscr{R}, d)$ forms a pseudometric space. Moreover, if the set of features $\{\mu_i\}$ is complete[1] in $\mathscr{R}$, then $(\mathscr{R}, d)$ forms a metric space.

[1] Depending upon which features are selected, the question of completeness may pose a problem of great difficulty. See, e.g., Kac [12].

It is also possible to select a set of features $\{\mu_\theta\}$ that are defined for $\theta$ in some interval, $I$. We assume then that the functions $\mu_\theta(R)$ can be embedded in some normed linear space of functions defined on $I$.

*Example. Convex regions; support functions.* Let $K$ be a bounded convex region in the $x$, $y$ plane. $K$ determines uniquely a $2\pi$-periodic continuous function $s_K(\theta)$, known as the *support function* of $K$. (See, e.g., Section 7).

We may take as our features $\mu_K(\theta) = s_K(\theta)$. If $L$ is a second convex region with support function $s_L(\theta)$, and if $\|\ \|$ designates a norm in the space of continuous, $2\pi$-periodic functions, then

$$d(K, L) = \| s_K(\theta) - s_L(\theta) \| \tag{5.9}$$

determines a metric on the set $\mathscr{K}$ of all bounded convex sets in the plane.

The use of the sup norm leads to the distance function

$$d(K, L) = \max_{0 \leq \theta \leq 2\pi} | s_K(\theta) - s_L(\theta) |. \tag{5.10}$$

It can be shown that this metric coincides with the Hausdorff metric. The function $d(K, L)$ defined by

$$d^2(K, L) = \int_0^{2\pi} (s_K(\theta) - s_L(\theta))^2 \, d\theta \tag{5.11}$$

is also a metric on $\mathscr{K}$ having a number of useful properties.

## 6. Best Approximation of Regions

Let $\mathscr{R}$ be a metric space of regions $R$. Let $\mathscr{R}^*$ be a subset of $\mathscr{R}$. For a given $R \in \mathscr{R}$ we may raise the question of finding an $R^* \in \mathscr{R}^*$ such that $d(R, R^*) = \inf_{R' \in \mathscr{R}^*} d(R, R')$. In general this is a nonlinear problem. If such an $R^*$ exists, then it solves the *problem of best approximation* of $R$ by a member of $\mathscr{R}^*$. In some happy cases, $R^*$ may both exist and be unique.

To construct examples of nonuniqueness: Consider the region $R$ that consists of two equal discs joined by a long thin corridor.

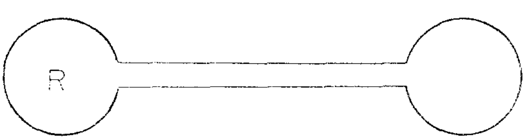


Figure 5.

Now, approximate this by a disc $R^*$ of equal radius in, say, the metric equal to the area of the symmetric difference.

Example 1. Given a fixed (nondegenerate) triangle $T$, what is the best approximation to it by a circle, using the area of the symmetric difference as a metric? We shall indicate the answer when the angles of $T$ are $<90^0$.

As observed before, there is a unique point $K$ in $T$, called *the Lemoine point* or the *symmedian point*, which minimizes the sums of squares of the distances to the sides of $T$. A basic property of the symmedian point is this. Through $K$ draw three lines that are antiparallel to the sides of $T$. These three lines intersect $\partial T$ in six points. These six points lie on a circle whose center is $K$. This circle is called the *cosine circle* of $T$ or the *second Lemoine circle*.[2] (See, e.g., Johnson [11, p. 271].)

The expression "antiparallel" is explained by the diagram below. The lines AB and CD are said to be *antiparallel* (with respect to the lines OA and OB) if $\sphericalangle$ ABO = $\sphericalangle$ OCD.

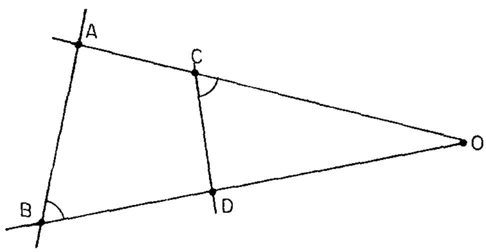


FIGURE 6.

The hexagon inscribed in the cosine circle is such that its opposite sides are equal and parallel. The *cosine circle solves the minimum problem*.[3] Here, in brief, are the reasons.

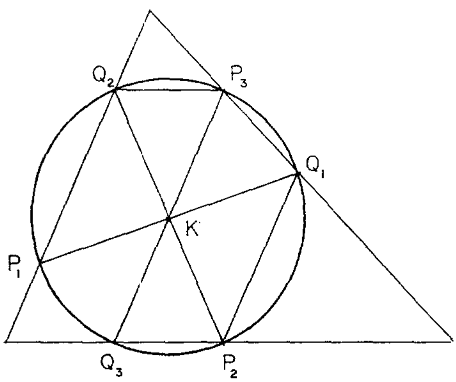


FIGURE 7.

[2] If $P_i : (x_i, y_i)$ are the vertices of $T$ and if $a_i$ designates the length of the side $P_i$, then the center $(x_L, y_L)$ and radius $r$ of the cosine circle are $x_L = (a_1^2x_1 + a_2^2x_2 + a_3^2x_3)/(a_1^2 + a_2^2 + a_3^2)$; similarly for $y_L$; $r = (a_1a_2a_2)/(a_1^2 + a_2^2 + a_3^2)$.

[3] This solution was very kindly supplied by Dr. Michael Goldberg of Washington, D.C.

An infinitesimal change in the radius will produce no change in the overlap area. Hence, this radius yields an extremal for the overlap area (in this case, a minimal overlap). Similarly, an infinitesimal translation of the center $K$ to the right will produce an increase in the overlaps $Q_2P_1$ and $P_1Q_3$, and an equal decrease in the overlaps $P_3Q_1$ and $Q_1P_2$. Similar translations parallel to the other sides of the triangle will produce no change in the overlap area. An infinitesimal translation in any other direction can be resolved into two translations parallel to sides of the triangle. Hence, the location $K$ yields an extremal for the overlap area. (These conditions are, in general, only necessary for an extremal; their sufficiency requires further arguments.)

EXAMPLE 2. We give another approximation of a triangle by a circle, using another feature. For the family of circles and triangles, the features $\tau_k(R) = \iint_R z^k\,dx\,dy$, $k = 0, 1, 2, 3$, form, as we have mentioned, a complete set. Let $T$ designate a fixed triangle whose vertices (in complex notation) are $z_1, z_2, z_3$. Let $C$ designate the disc centered at $z = w$ with radius $r > 0$. Introduce

$$d^2(T, C) = \sum_{k=0}^{3} |\tau_k(T) - \tau_k(C)|^2. \tag{6.1}$$

By the complex mean value theorem which says that $\iint_C f(z)\,dx\,dy = \pi r^2 f(w)$ for $f(z)$ analytic in $C$, it follows that $\tau_k(C) = \pi r^2 w^k$. The values of $\tau_k(T)$ are as follows (see Davis [5])

$\tau_0(T) = A$ = area of $T$; $\tau_1(T) = (A/3)\,s$; $\tau_2(T) = (A/6)(s^2 - t)$, $\tau_3(T) = (A/10)(s^3 - 2st + 7p)$, where $s = z_1 + z_2 + z_3$, $t = z_1z_2 + z_2z_3 + z_3z_1$, $p = z_1z_2z_3$.

Hence

$$\begin{aligned} d^2(T, C) = {} & |A - \pi r^2|^2 + |\tau_1 - \pi r^2 w|^2 + |\tau_2 - \pi r^2 w^2|^2 \\ & + |\tau_3 - \pi r^2 w^3|^2. \end{aligned} \tag{6.2}$$

Letting $c_0 = 1/\pi$, $c_i = \tau_i/A\pi$, $i = 1, 2, 3$, $\sigma = r^2/A$, one has

$$\frac{d^2(T, C)}{(A\pi)^2} = |c_0 - \sigma|^2 + |c_1 - \sigma w|^2 + |c_2 - \sigma w^2|^2 + |c_3 - \sigma w^3|^2. \tag{6.3}$$

This is to be minimized over the three-dimensional set $|w| < \infty$, $\sigma > 0$, a semilinear problem.

We have

$$\frac{d^2(T, C)}{(A\pi)^2} = \sigma^2(1 + |w|^2 + |w|^4 + |w|^6) - 2\sigma\left(\mathrm{Re}\sum_{k=0}^{3} \bar{c}_k w^k\right) + \sum_{k=0}^{3} |c_k|^2. \tag{6.4}$$

Let

$$p(w) = \sum_{k=0}^{3} \bar{c}_k w^k. \tag{6.5}$$

Now let $w$ be fixed. If Re $p(w) < 0$, then $d^2$ is minimized at $\sigma = 0$ and its minimum value is $(A\pi)^2 \sum_{k=0}^{3} | c_k |^2$. If Re $p(w) \geqslant 0$, then $d^2$ is minimized at $\sigma = (\mathrm{Re}\, p(w))/(1 + | w |^2 + | w |^4 + | w |^6)$ and its minimum value is

$$(A\pi)^2 \left( \sum_{k=0}^{3} | c_k |^2 - \frac{(\mathrm{Re}\, p(w))^2}{1 + | w |^2 + \cdots + | w |^6} \right).$$

The investigation should now be carried further by varying $w$.

As a simple case, take $T$ as the equilateral triangle whose vertices are $z_1 = 1$, $z_2 = \omega$, $z_3 = \omega^2$; $\omega^3 = 1$, $\omega \neq 1$. Then, $s = 0$, $t = 0$, $p = 1$, $p(z) = (1/\pi)(1 + (7/10)\, z^3)$. It can be shown that, as expected, the minimizing circle has center at $w = 0$ and area equal to that of $T$.

## 7. Approximation in the Hausdorff Metric

In this section we consider the approximation problem where the measure of discrepancy between two sets is given by the Hausdorff distance. A natural setting is the family $\mathscr{C}$ of nonempty, compact subsets of the plane $R^2$. We recall that if $C_1, C_2 \in \mathscr{C}$, then the Hausdorff distance between them is given by

$$d(C_1, C_2) = \min\{\epsilon > 0 \mid C_1 \subseteq C_2 + \epsilon\bar{0},\ C_2 \subseteq C_1 + \epsilon\bar{0}\}, \tag{7.1}$$

where $\bar{0}$ is the closed unit disc in the plane under the Euclidean metric. Here $A + B$ designates the Minkowski sum i.e., the set of all $a + b$ with $a \in A$, $b \in B$. The scalar product $\epsilon A$ designates the set of all $\epsilon a$ with $a \in A$. Under the Hausdorff metric, $\mathscr{C}$ is a complete metric space (see, for instance, [17]). By considering sets of rational points, we see that $\mathscr{C}$ is separable. Moreover, any subfamily $\{C \in \mathscr{C} \mid C \subseteq C_0,\ C_0 \in \mathscr{C} \text{ fixed}\}$ is compact. This follows by noting that any such subfamily is both closed and totally bounded (an $\epsilon$-net can be formed from a sufficiently fine lattice). For convex sets, this compactness result is due to Blaschke (see [27]).

With this preparation, we can use standard techniques to investigate best approximation.

THEOREM. *Let $A \in \mathscr{C}$ and let $\mathscr{B} \subseteq \mathscr{C}$ be a closed subfamily. Then for some $B_0 \in \mathscr{B}$*

$$d(A, B_0) = \inf\{d(A, B)\ B \in \mathscr{B}\}. \tag{7.2}$$

*Proof.* Fix any $\tilde{B} \in \mathscr{B}$ and define

$$\tilde{\mathscr{B}} = \{B \in \mathscr{B} \mid d(A, B) \leqslant d(A, \tilde{B})\}$$

Clearly,

$$\inf\{d(A, B) | B \in \mathscr{B}\} = \inf\{d(A, B) | B \in \tilde{\mathscr{B}}\}.$$

Moreover, $\tilde{\mathscr{B}}$ is compact since it is both closed and a subfamily of the compact $\{C \in \mathscr{C} \mid C \subseteq A + d(A, \tilde{B})\,\bar{0}\}$. Hence, $d(A, B)$ attains a minimum over $\tilde{\mathscr{B}}$.

Uniqueness of a best approximating set cannot generally be asserted. Consider, for instance, $A =$ the closed unit disc, $\mathscr{B} =$ the set of line segments $[-b, +b]$. The minimum Hausdorff distance is equal to one and is attained for every $b \in [0, 2]$. This example, however, suggests the following characterization of the subfamily of best approximants.

THEOREM. *Let $A \in \mathscr{C}$ and let $\mathscr{B} \subseteq \mathscr{C}$ be closed. Then the subfamily $\mathscr{B}_0 \subseteq \mathscr{B}$ of best approximants is compact.*

*Proof.* Using the notation of the previous theorem, we have

$$\mathscr{B}_0 = \tilde{\mathscr{B}} \cap \{C \in \mathscr{C} \mid d(A, C) = d(A, B_0)\}.$$

Since $\tilde{\mathscr{B}}$ is compact and $\{C \in \mathscr{C} \mid d(A, C) = d(A, B_0)\}$ is closed, $\mathscr{B}_0$ is compact.

By suitably restricting the family of sets under consideration, stronger results may be available. An example is *affine approximation*, where it is required to approximate a set $A$ from the family of translations and scalings of a fixed set $B$. With the assumption of convexity, uniqueness can be asserted.

THEOREM. *Let $A, B \in \mathscr{C}$ be convex with $B$ satisfying*

(i) *$B$ has an interior point,*

(ii) *$\partial B$ has no corners.*

*Let $\mathscr{B} = \{rB + p \mid r \geqslant 0, p \in R^2\}$.*
*Then* $\inf\{d(A, \tilde{B}) | \tilde{B} \in \mathscr{B}\}$ *is attained uniquely for some $B_0 \in \mathscr{B}$. Moreover,*

$$d(A, B_0) = \inf\{\epsilon > 0 \mid A \subseteq B_0 + \epsilon\bar{0}\} \tag{7.3}$$
$$= \inf\{\epsilon > 0 \mid B_0 \subseteq A + \epsilon\bar{0}\}. \tag{7.4}$$

Before proceeding to the proof of the theorem, we present some facts about convex sets which can be found in standard references such as Eggleston [8], Yaglom and Boltyanskii [27], Rockafellar [18], and Valentine [22]. Associated uniquely to each convex $K \in \mathscr{C}$ is its continuous $2\pi$-periodic *support function*

$$s_K(\theta) = \max\{x \cos\theta + y \sin\theta \mid (x, y) \in K\}.$$

This association preserves both metric and "semilinear" structure

$$s_{\alpha K_1+\beta K_2}(\theta) = \alpha s_{K_1}(\theta) + \beta s_{K_2}(\theta), \qquad \alpha, \beta \geqslant 0, \tag{7.5}$$

$$d(K_1, K_2) = \| s_{K_1} - s_{K_2} \| = \max\{| s_{K_1}(\theta) - s_{K_2}(\theta)| \mid \theta \in [0, 2\pi)\}. \tag{7.6}$$

In addition,

$$\inf\{\epsilon > 0 \mid K_1 \subseteq K_2 + \epsilon\bar{0}\} = \max\{s_{K_1}(\theta) - s_{K_2}(\theta) \mid \theta \in [0, 2\pi)\}. \tag{7.7}$$

Finally, if $K$ contains the origin, then

$$s_K(\theta) \geqslant 0, \qquad \text{for all } \theta, \tag{7.8}$$

with strict inequality if the origin is an interior point.

*Proof of the Theorem.* We begin by interpreting the setup in terms of support functions. Let $s_A(\theta)$ be the support function of $A$. By translating $A$ appropriately, we can assume without loss of generality that $A$ contains the origin and hence $s_A(\theta) \geqslant 0$. Let $s_B(\theta)$ be the support function of $B$, where again, through a suitable translation, the origin is an interior point and $s_B(\theta) > 0$.

If $p = (x_p, y_p)$, then the support function of $rB + p$, $r \geqslant 0$, is $rs_B(\theta) + x_p \cos\theta + y_p \sin\theta$, and

$$d(A, rB + p) = \max\{| s_A(\theta) - (rs_B(\theta) + x_p \cos\theta + y_p \sin\theta)| \mid \theta \in [0, 2\pi)\}. \tag{7.9}$$

We seek to minimize this expression over all values $r \geqslant 0$, $x_p$, $y_p$. For the moment, we relax the requirement that $r$ be nonnegative and recognize the resulting problem as one of best uniform approximation of $s_A(\theta)$ by a linear combination of the functions $s_B(\theta)$, $\cos\theta$, $\sin\theta$.

It is well known that if these functions form a periodic Tchebycheff system, then a unique best approximation exists (see, for example, Karlin and Studden [13, p. 282]). This condition holds if

$$\begin{vmatrix} s_B(\theta_1) & \cos\theta_1 & \sin\theta_1 \\ s_B(\theta_2) & \cos\theta_2 & \sin\theta_2 \\ s_B(\theta_3) & \cos\theta_3 & \sin\theta_3 \end{vmatrix} > 0 \tag{7.10}$$

or, equivalently,

$$s_B(\theta_1)\sin(\theta_3 - \theta_2) + s_B(\theta_2)\sin(\theta_1 - \theta_3) + s_B(\theta_3)\sin(\theta_2 - \theta_1) > 0 \tag{7.11}$$

where $\theta_1, \theta_2, \theta_3$ are distinct and taken in counterclockwise order [13, p. 180].

There are two possibilities. If $\theta_1$, $\theta_2$, $\theta_3$ are "spread out," i.e., $0 < \theta_2 - \theta_1 \leqslant \pi$, $0 < \theta_3 - \theta_2 < \pi$, $0 < \theta_1 - \theta_3 < \pi$, then each term in (7.11) is nonnegative and two must be strictly positive (recall $s_B(\theta) > 0$, by assumption). On the other hand, if the $\theta$'s are "clustered," $\theta_1 < \theta_2 < \theta_3 < \theta_1 + \pi$, a more delicate analysis is required. It is well known that (7.11) always holds with $>$ replaced by $\geqslant$ (see, for instance Boas [3, p. 70], Vitale [25]). If equality obtains, then the following type of picture must occur.

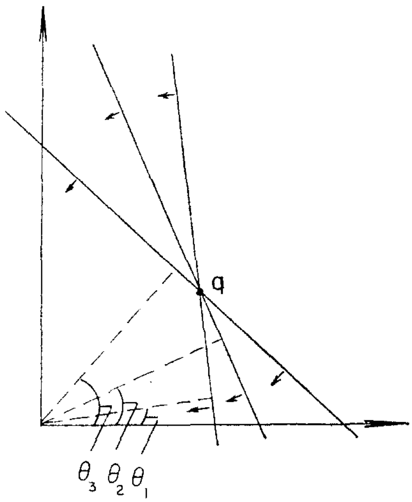


FIGURE 8.

*B* must lie in the intersection of the depicted half planes *and* it must contain the point $q$. Necessarily, $q$ is a corner of $\partial \boldsymbol{B}$, but this contradicts assumption (ii).

Hence we have verified (7.11) and conclude that there is a best approximating linear combination, which we write as $rs_B(\theta) + a\cos(\theta - \theta_0)$. Indeed, the theory allows more to be concluded. If

$$M = \max\{|s_A(\theta) - (rs_B(\theta) + a\cos(\theta - \theta_0))| : \theta \in [0, 2\pi]\} \tag{7.12}$$

then there exist four points of equioscillation, namely, $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$ (again in counterclockwise order) such that

$$M = (-1)^i\, \delta[s_A(\theta_i) - (rs_B(\theta_i) + a\cos(\theta_i - \theta_0))]$$

where $i = 1, 2, 3, 4$ and $\delta = +1$ or $-1$. This implies (7.3) and (7.4) and to conclude the proof we only have to show that $r > 0$. Using (7.12), we have

$$M \geqslant s_A(\theta_0 + \pi) - (rs_B(\theta_0 + \pi) - a)$$

or

$$a - M \leqslant rs_B(\theta_0 + \pi) - s_A(\theta_0 + \pi) \leqslant D \tag{7.13}$$

where

$$D = \max\{rs_B(\theta) - s_A(\theta) |\ \theta \in [0, 2\pi)\}.$$

Again from (7.12) we have

$$M \leqslant \max\{| s_A(\theta) - rs_B(\theta)| |\ \theta \in [0, 2\pi)\} + \max\{| a \cos(\theta - \theta_0)| |\ \theta \in [0, 2\pi)\}$$
$$= D + a$$

or

$$(M - a) \leqslant D. \tag{7.14}$$

Combining (7.13) and (7.14) yields $D \geqslant | M - a | \geqslant 0$ and the continuity of $s_A$ and $s_B$ assures the existence of a $\tilde{\theta}$ such that

$$D = rs_B(\tilde{\theta}) - s_A(\tilde{\theta}) \geqslant 0$$

so that

$$rs_B(\tilde{\theta}) \geqslant s_A(\tilde{\theta}) \geqslant 0.$$

Since $s_B(\tilde{\theta}) > 0$, it follows that $r \geqslant 0$.

The actual numerical construction of the best approximating set would proceed through the use of generalized Remez algorithms. We turn now to a very special case in which this optimal set can be described in simple terms. It is required to find the best approximating disc to a given triangle $T$. Some notation will be useful. The center and radius of a disc $C$ will be denoted by $p$ and $r$, respectively. Let $M$ denote the maximum distance from $p$ to a vertex of $T$ and let $m$ denote the minimum distance between $p$ and a point of $\partial T$. If $p \notin T$, then, setting $x_+ = x$ for $x \geqslant 0$, $x_+ = 0$, otherwise,

$$d(C, T) = \max\{(M - r)_+ , m + r\}$$

which is minimized for $r = (M - m)/2$, yielding $d(C, T) = (m + M)/2$. Clearly, by moving $p$ "toward" $T$, we can reduce $(m + M)/2$. Hence, a candidate for the best approximating disc cannot have its center outside $T$. For discs centered at $p \in T$, the optimal radius is $(M + m)/2$, yielding $d(C, T) = (M - m)/2$. Hence, in searching for the best disc, it is sufficient to minimize $M - m$ over all choices of centers $p \in T$.

THEOREM. *Given any triangle, the best approximating disc is unique. Its center is the intersection of the bisector of the smallest angle and the perpendicular bisector of the largest side. (The case of isosceles triangles should be interpreted appropriately.)*

*Proof.* Uniqueness follows from the previous theorem. By the preceding remarks, we only have to verify that the described point minimizes $M - m$. We separate into two cases.

*Case* 1. The largest angle of the triangle $\geq 90°$. We draw the triangle with its largest side $L$ down and its smallest angle to the right.

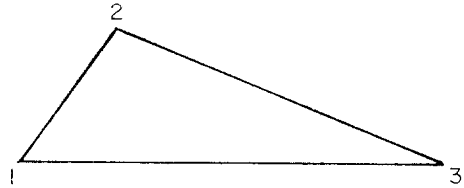


FIGURE 9.

Let $p$ be any point of the triangle. Note that a farthest vertex from $p$ must be 1 or 3 (possibly both). Suppose without loss of generality it is 3. If $p$ is not closest to $L$, rotate $p$ around vertex 3 to get a point $p'$ with the same $M$ and $m$, and hence $M - m$, as $p$.

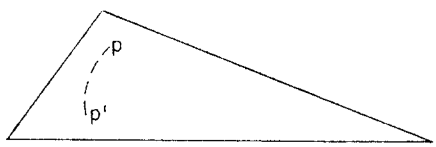


FIGURE 10.

Now construct $p''$ by projecting $p'$ onto the perpendicular bisector of $L$.

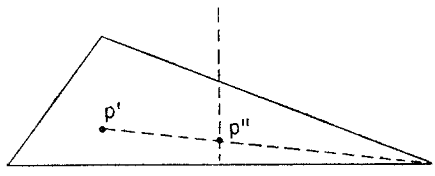


FIGURE 11.

It is straightforward to verify that $p''$, if $\neq p'$, is strictly better than $p'$, i.e. (using an obvious notation) $M'' - m'' < M - m$. Thus the center of the best

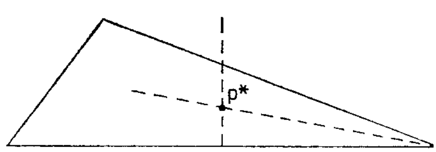


FIGURE 12.

circle must lie on the perpendicular bisector. Let $p^*$ be the point described in the theorem. Note that the points above $p^*$ on the perpendicular bisector have strictly larger $M$ and smaller $m$ than $p$. As for points below $p^*$, we have the blown-up picture (where with a slight abuse of notation, we denote by $L/2$ half the length of side $L$).

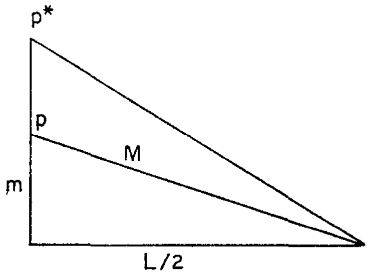


FIGURE 13.

Note that

$$M - m = (m^2 + (L^2/4))^{1/2} - m.$$

Since

$$\frac{d(M-m)}{dm} = \frac{m}{(m^2 + (L^2/4))^{1/2}} - 1 < 0,$$

$m$ should be made as large as possible.

Hence $p^*$ is the best center.

*Case* 2. The triangle is acute angled. The perpendicular bisectors divide the triangle into three regions labeled by the farthest vertex.

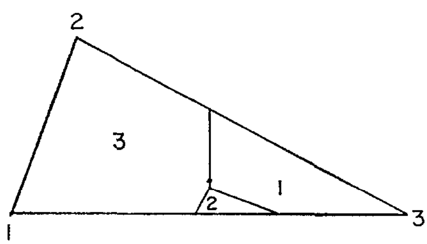


FIGURE 14.

By an argument similar to that used in Case 1, it is possible to show that $p^*$ must lie on the "skeleton" and indeed must coincide with the intersection of the upper branch and the angle bisector at vertex 3.

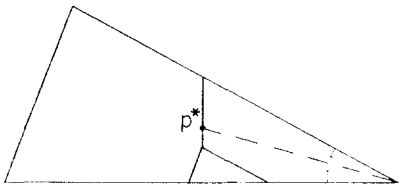


FIGURE 15.

## 8. APPROXIMATION AS A PROJECTION OPERATOR

The interpretation of least-squares approximation as a projection is well known. There has been considerable abstract development of this idea, particularly for functions of one real variable. See, e.g., Deutsch [6].

Let us suppose that $\mathscr{R}$ is a family of regions in the plane and we have a mapping $P$ of $\mathscr{R}$ into $\mathscr{R}$. Such a mapping will be called a *projection* if $P^2 = P$.

EXAMPLE 1. Let $\mathscr{R}$ consist of all (nondegenerate) triangles and circles in the plane. If $T$ is a triangle, let $P(T)$ designate the nine-point circle (or the inscribed circle or the cosine circle,...) of $T$; fix on one of these. If $C$ is a circle, let $P(C) = C$. Then, clearly $P^2 = P$ so that $P$ is a projection.

EXAMPLE 2. Let $\mathscr{K}$ designate the set of all bounded convex regions $K$ of the plane. Let $P(K)$ designate the circumscribing polygon with sides in given feasible directions (where the directions have been specified by giving, e.g., the direction of the outward normals to the sides). Then, clearly, $P^2 = P$.

EXAMPLE 3. Let $\mathscr{R}$ be a metric space of regions and $\mathscr{R}^*$ a subset of $\mathscr{R}$. Suppose, further, that for each $R \in \mathscr{R}$ the problem of best approximation to $R$ from among the regions of $\mathscr{R}^*$ has a *unique solution*. Designate it by $P_{\mathscr{R}^*}(R)$. Clearly, $P_{\mathscr{R}^*}$ is a projection operator.

EXAMPLE 4. Let $\mathscr{K}$ designate the set of all compact convex regions $K$ of the plane. Let $P(K)$ designate the unique disc which best approximates $K$ in the sense of the Hausdorff metric. Then $P^2 = P$.

The study of the properties of projection operators in the present context would seem to be worthwhile.

## 9. Best Approximation of Regions: A Stochastic Approach

If the problem is of sufficient complexity and numerical answers are desired, we might be forced to a Monte Carlo approach. Take, for example, the problem of approximating a fixed triangle by a circle in the area of the symmetric difference metric. Let the characteristic function of $T$ be $C_T(x, y)$ while that of a circle centered at $(\alpha, \beta)$ and of radius $r$ be $C_{\alpha,\beta,r}(x, y)$. We must minimize

$$J(\alpha, \beta, r) = \iint_{-\infty}^{\infty} | C_T(x, y) - C_{\alpha,\beta,r}(x, y)| \, dx \, dy \tag{9.1}$$

over the region $-\infty < \alpha, \beta < \infty, r > 0$. Insofar as the boundary of the symmetric difference is a sufficiently complicated piecewise linear and circular configuration, it might be considered the better part of wisdom to estimate $J(\alpha, \beta, r)$ by Monte Carlo techniques.

The naive approach would be as follows. (1) Fix $\alpha, \beta, r$ and then estimate $J(\alpha, \beta, r)$ by Monte Carlo to "sufficient" accuracy. Then, (2) vary $\alpha$, $\beta$, and $r$ according to some strategy of minimization, e.g., a gradient method.

It might occur to the reader that it should be possible to combine (1) and (2) into one process, varying the parameters as one samples the integrand.

Such a process has been called "*stochastic approximation*" and extensive information about it can be found in Wasan [26, Chap. 3]. While convergence theorems have been proved, the hypotheses under which the theorems hold are often "unverifiable," depending, as they must, upon the nature of $J$ as a function of its parameters, and the nature of the minimizing and iterative strategies. Furthermore, even if one were to exhibit convergence in one's particular problem, it is by no means clear that the method of "stochastic approximation" exhibits computational economies over the naive method. Numerical experimentation with techniques of stochastic approximation have revealed that convergence may be agonizingly slow.

## 10. Stochastic Approximation of a Second Kind

We shall begin with a very specific problem in geometrical probability. Let $S$ designate the unit square $S$: $0 \leqslant x, y \leqslant 1$. Let $T$ designate a closed (interior plus boundary) triangle contained in $S$. If $T_1$ and $T_2$ are two such triangles selected at random, *what is the probability that $T_1$ and $T_2$ overlap*? More specifically, a random triangle $T$ will be constructed by selecting independently six numbers $x_i$, $y_i$, $i = 1, 2, 3$, from a uniform distribution on $0 \leqslant t \leqslant 1$ and using $(x_i, y_i)$, $i = 1, 2, 3$, as the vertices of the triangle.

In the language of linear programming, consider the system of six linear inequalities

$$A_i x + B_i y + C_i \geqslant 0, \qquad i = 1, 2, ..., 6. \tag{10.1}$$

What is the probability that the system is *feasible*, given certain information about the distribution of the $A_i$, $B_i$, and $C_i$?

For two given triangles (or six given linear inequalities), algorithms of linear programming may be used to determine numerically whether the triangles do or do not overlap.

To solve such a problem may involve a considerable amount of computation. Therefore we might like to replace this problem by a simpler one, but which is only partially equivalent to it: Approximate each triangle by a circle (closed disc) by means of a fixed policy of approximation. Then ask the question: Do the corresponding circles overlap? This can be answered by a short computation.

As examples of a fixed policy of approximation of a triangle $T$ by a circle $C$ we mention:

(a) use of the circle whose center is at the center of gravity of $T$ and whose area equals that of $T$.

(b) use of the circle whose center is at the center of gravity $M$ of $T$ and whose radius equals the average distance of $M$ from the vertices of $T$.

(c) use of second Lemoine circle of $T$.

These particular policies are coordinate free.

A policy of approximation might be completely deterministic. But it also might be stochastic or might have a mixture of deterministic and stochastic elements. For example, given $T$, determine $C$ by using the center of gravity of $T$ as its center and by selecting its radius at random from a uniform distribution on $0 \leqslant r \leqslant 1$.

In what follows we shall assume that the policy is deterministic.

For each fixed policy $\mathscr{P}$ of approximation, we consider the $2 \times 2$ matrix $P$ whose elements $p_{ij}$ are:

$p_{11} =$ the probability that both the triangles and the corresponding circles overlap;

$p_{12} =$ the probability that the triangles overlap but the corresponding circles do not;

$p_{21} =$ the probability that the triangles do not overlap but the corresponding circles do;

$p_{22} =$ the probability that neither the triangles nor the corresponding circles overlap.

Notice that

$$p_{11} + p_{12} = \text{the probability that the triangles overlap;}$$

$$p_{11} + p_{22} = p = \text{trace } P = \text{the probability that the behavior of the circles predicts properly the behavior of the triangles.} \tag{10.2}$$

For each policy of approximation $\mathscr{P}$ we may, in principle, compute the corresponding $p = p(\mathscr{P})$. For a given family of policies $\{\mathscr{P}\}$, we may raise the questions of whether there is a *best* one, how to characterize it, how to compute it, etc.

We shall call this type of approximation *stochastic approximation.* This term is used in a different context within sampling theory (see Section 9); however, we feel that this term is equally, if not more, appropriate to the process just mentioned.

*Numerical Values by Monte Carlo*

The basic probabilities $p_{ij}$ where estimated by Monte Carlo in the following way. Each of the 12 coordinates was obtained from the FORTRAN random number generator and was assumed to be drawn from a uniform distribution over [0, 1]. The policy $\mathscr{P}_1$ for the approximating circle was to place its center at the center of gravity $M$ of the triangle and to use the average distance of $M$ to the vertices as its radius. The results are tabulated below.

Key

| | |
|---|---|
| Triangles overlap<br>Circles overlap | Triangles overlap<br>Circles do not overlap |
| Triangles do not overlap<br>Circles overlap | Neither triangles nor<br>Circles overlap |

Number of runs: $n = 10{,}000$

| | |
|---|---|
| 62.92 % | 0.16 % |
| 27.39 % | 9.53 % |

Number of runs: $n = 20{,}000$

| | |
|---|---|
| 63.43 % | 0.16 % |
| 26.915 % | 9.475 % |

Adopting the values after $n = 20{,}000$ experiments, one can say that the probability that two triangles overlap is $0.6345 + 0.0016 \approx 0.64$. The probability that the circles are an accurate predictor for the triangles is $0.6345 + 0.09475 \approx 0.73 = p(\mathscr{P}_1)$.

The upper right element of the matrix is particularly interesting geometrically, showing a relatively rare event. Consider also two degenerate policies.

(1) A "placebo" policy $\mathscr{P}_u$ of approximating $T$ by a fixed circle, say the unit circle. Since the approximating circles always overlap, this policy will be effective $\approx 64\%$ of the time.

(2) The "null" policy $\mathscr{P}_n$ of approximating $T$ by a null circle of radius 0 and center at the c.g. of $T$. Since these points coincide with probability 0, this policy will be effective $\approx 36\%$ of the time.

We list here the results of several additional policies $\mathscr{P}_2, \mathscr{P}_3, \mathscr{P}_4, \mathscr{P}_5, \mathscr{P}_6$.

$\mathscr{P}_2$: center of circle is at c.g. $M$ of $T$. Radius of circle equals maximum distance from $M$ to vertices of $T$.

$\mathscr{P}_3$: center of circle is at c.g. $M$ of $T$. Radius of circle equals minimum distance from $M$ to vertices of $T$.

$\mathscr{P}_4$: center of circle is at c.g. $M$ of $T$. Radius of circle equals the minimum of the distances of $M$ to three sides of $T$.

$\mathscr{P}_5$: Center of circle is at c.g. of $T$. Area of circle is area of $T$.

$\mathscr{P}_6$: Circle is the second Lemoine circle of $T$.

The first four matrices below are all for 20,000 samples.

$$P(\mathscr{P}_2)$$

$$\begin{matrix} 63.61\% & 0\% \\ 33.535\% & 2.855\% \end{matrix}$$

$$p(\mathscr{P}_2) = 0.665$$

$$P(\mathscr{P}_3)$$

$$\begin{matrix} 55.19\% & 8.42\% \\ 9.615\% & 26.775\% \end{matrix}$$

$$p(\mathscr{P}_3) = 0.8197$$

$$P(\mathscr{P}_4)$$

$$\begin{matrix} 18.33\% & 45.28\% \\ 0\% & 36.39\% \end{matrix}$$

$$p(\mathscr{P}_4) = 0.547$$

$$P(\mathscr{P}_5)$$

$$\begin{matrix} 45.88\% & 17.73\% \\ 3.12\% & 33.27\% \end{matrix}$$

$$p(\mathscr{P}_5) = 0.792$$

$P(\mathscr{P}_6)$

| | |
|---|---|
| 49.45 % | 27.92 % |
| 0.32 % | 22.30 % |

$$p(\mathscr{P}_6) = 0.718$$

In $\mathscr{P}_6$, the number of samples: 4000; is restricted to acute-angled triangles.

Thus, of the five policies tested, $\mathscr{P}_3$ appears to be optimal.

We may run statistical tests on these results to determine the level of confidence which we can place in distinguishing, say, $\mathscr{P}_3$ from $\mathscr{P}_5$.

A commonly employed test is as follows. Assume that an event has occurred $a$ times in $m$ repetitions of condition (1) and $b$ times in $n$ repetitions of condition (2). Then, the difference in relative frequencies is regarded as significant if

$$\left|\frac{a}{m} - \frac{b}{n}\right| > \frac{1}{2}\left(\frac{1}{m} + \frac{1}{n}\right) + u_\alpha \left(\frac{(a+b)(m+n-a-b)}{mn(m+n)}\right)^{1/2}, \qquad (10.3)$$

where $u_\alpha$ is the value of the standard normal distribution at the significance level $\alpha$. (That is, for $\alpha = 95\%$, $u_\alpha = 1.96$; for $\alpha = 99\%$, $u_\alpha = 2.576$.)

According to this test, the difference between $\mathscr{P}_3$ and $\mathscr{P}_5$ checks out as significant at better than the 99.99 % level of significance.

*Formalization of the Problem*

Let the vertices of $T_1$ be $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$ and those of $T_2$ be $(x_4, y_4)$, $(x_5, y_5)$, $(x_6, y_6)$. Let $v$ designate the 12-component vector $x_1$, $y_1, \ldots, x_6, y_6$. Let $H_{12}$ designate the closed unit hypercube in (real) Euclidean 12-space.

Define a characteristic function $\phi(v)$ on $H_{12}$ as

$$\begin{aligned} \phi(v) &= 1 \quad \text{if} \quad T_1 \cap T_2 \neq 0, \\ \phi(v) &= 0 \quad \text{if} \quad T_1 \cap T_2 = 0. \end{aligned} \qquad (10.4)$$

Thus, $\phi$ is 1 on a certain polyhedron lying in $H_{12}$, and 0 elsewhere. Now one has,

$$\int_{H_{12}} \phi(v)\, dv \qquad (dv = dx_1\, dx_2 \cdots dy_5\, dy_6) \qquad (10.5)$$

= the probability that the two triangles overlap.

Fix a policy $\mathscr{P}$ of approximation of a triangle $T$ by a circle $C$.

A given vector $v$ determines two triangles $T_1$ and $T_2$, which, in turn, using the policy $\mathscr{P}$, determine two approximating circles $C_1$ and $C_2$.

Define a second function $\psi(v)$ as

$$\begin{aligned} \psi(v) &= 1 \qquad \text{if} \quad C_1 \cap C_2 \neq 0, \\ \psi(v) &= 0 \qquad \text{if} \quad C_1 \cap C_2 = 0. \end{aligned} \tag{10.6}$$

The policy $\mathscr{P}$ will be a good one if $\psi$ is close to $\phi$, meaning by this that $\|\phi - \psi\|$ is small, where $\|\ \|$ designates a norm on some class of functions defined on $H_{12}$. For simplicity, adopt $\|f\|^2 = \int_{H_{12}} f^2\,dv$, so that one wants a small

$$I(\mathscr{P}) = \int_{H_{12}} (\phi - \psi)^2\,dv. \tag{10.7}$$

Notice that whenever the circles are a good predictor, $\phi(v) = \psi(v)$, so that $(\phi - \psi)^2 = 0$. If the circles are a bad predictor, $\phi - \psi = \pm 1$ so that $(\phi - \psi)^2 = 1$. Hence

$$I(\mathscr{P}) = \text{the probability that the circles are a bad predictor.} \tag{10.8}$$

Naturally we would like to minimize it.

Insofar as $\phi$ and $\psi$ take on only the values 0 and 1, $\phi = \phi^2$, $\psi = \psi^2$, and we note the alternative expression

$$I(\mathscr{P}) = \int_{H_{12}} (\phi + \psi - 2\phi\psi)\,dv. \tag{10.9}$$

## 4. Approximation Theory

A given policy $\mathscr{P}$ of approximation determines the function $\psi$ on $H_{12}$. Thus, through the correspondence $\mathscr{P} \to \psi$, and using $I(\mathscr{P})$, we have converted our problem to one of classical approximation theory.

If the family of policies $\{\mathscr{P}_\lambda\}$ consists only of a finite number of distinct policies, $\mathscr{P}_1, \ldots, \mathscr{P}_N$, then from the analytical point of view there is nothing further to discuss. There is an optimal policy and the question of its expeditious computation is another matter.

If the family of policies $\{\mathscr{P}_\lambda\}$ is infinite, then there is a theory to be developed, and one must look at the corresponding family of approximants $\{\psi_\lambda(v)\}$ where $\mathscr{P}_\lambda \to \psi_\lambda$.

In general, the subject of optimal policy is a nonlinear problem. Existence of a best approximation is usually based upon a compactness argument and uniqueness of best approximation can often, but not always, be based upon a convexity argument.

Let $T_1$ have vertices $(x_i, y_i)$, $i = 1, 2, 3$, while $T_2$ has vertices $(x_i', y_i')$.

$i = 4, 5, 6$. Let the deterministic policy $\mathscr{P}$ assign to $T_1$ the circle centered at $(a, b)$ and of radius $r_1$ where one assumes that

$$\begin{aligned} a &= q(x_1, \ldots, y_3), \\ b &= s(x_1, \ldots, y_3), \\ r_1 &= r(x_1, \ldots, y_3). \end{aligned} \tag{10.10}$$

The policy $\mathscr{P}$ assigns to $T_2$ the circle centered at $(c, d)$ and of radius $r_2$ where

$$\begin{aligned} c &= q(x_4, \ldots, y_6), \\ d &= s(x_4, \ldots, y_6), \\ r_2 &= r(x_4, \ldots, y_6). \end{aligned} \tag{10.11}$$

These two circles will be disjoint if and only if

$$((a - c)^2 + (b - d)^2)^{1/2} > r_1 + r_2 . \tag{10.12}$$

Hence, writing

$$\begin{aligned} \Omega(v) &= ((a - c)^2 + (b - d)^2)^{1/2}, \\ \omega(v) &= r_1 + r_2 , \end{aligned} \tag{10.13}$$

the circles are disjoint if and only if $\Omega(v) > \omega(v)$. Hence,

$$\begin{aligned} \psi(v) &= 1 \qquad \text{if} \quad \Omega(v) \leqslant \omega(v), \\ \psi(v) &= 0 \qquad \text{if} \quad \Omega(v) > \omega(v). \end{aligned} \tag{10.14}$$

Now let $\{\mathscr{P}_\alpha\}$ designate a family of policies which are parametrized by $\alpha$, where we assume that $\alpha$ is a real variable or a vector of real variables. Each $\mathscr{P}_\alpha$ determines two families of functions, $\Omega(\alpha; v)$ and $\omega(\alpha; v)$. Through them, one has

$$\begin{aligned} \psi_\alpha(v) &= 1 \qquad \text{if} \quad \Omega(\alpha; v) \leqslant \omega(\alpha; v), \\ \psi_\alpha(v) &= 0 \qquad \text{if} \quad \Omega(\alpha; v) > \omega(\alpha; v). \end{aligned} \tag{10.15}$$

Example. Let the vertices of $T$ be $A$, $B$, $C$, and its center of gravity be $M$. Let $\rho_1 = \overline{AM}$, $\rho_2 = \overline{BM}$, $\rho_3 = \overline{CM}$. Let the circle $C$ approximating $T$ have its center at $M$ and have radius

$$r = (\tfrac{1}{3}(\rho_1{}^\alpha + \rho_2{}^\alpha + \rho_3{}^\alpha))^{1/\alpha},$$

$\alpha > 0$ fixed.

We now assume,

1. $\alpha$ varies on a compact subset $A$ of the $\alpha$ space.

2. the two functions $\Omega(\alpha; v)$ and $\omega(\alpha; v)$ are continuous for $\alpha \in A$, $v \in H_{12}$.

Under these conditions, a minimum to $I$ exists. For.

$$\begin{aligned} I(\mathscr{P}_\alpha) = I(\alpha) &= \int_{H_{12}} (\phi(v) - \psi_\alpha(v))^2 \, dv \\ &= \int_{H_{12}} \phi(v)\, dv + \int_{\Omega(\alpha, v) \leqslant \omega(\alpha; v)} (1 - 2\phi(v))\, dv. \end{aligned} \tag{10.16}$$

In view of the continuity of $\Omega$ and $\omega$, it is clear from standard theorems of integration theory that the second integral is continuous in $\alpha$ so that $I(\alpha)$ is continuous. The existence of a minimum is an immediate consequence.

*Adaptive learning.* We may set the problem of having a computing machine teach itself what is the best policy $\mathscr{P}$ of approximation. One assumes that the policy space has been parameterized in some fashion. Insofar as theories of adaptive learning are clearly related to stochastic approximation in the sense of Section 9, it follows that both types of stochastic approximation can, in fact, be interrelated.

## Bibliographical Remarks

Professor I. J. Schoenberg has kindly pointed out that the affine approximation theorem (Section 7) for the special case of $B$, a disc, has been considered by Lebesgue (Sur quelques questions de minimum, relatives aux courbes orbiforms, et sur leur rapports avec le calcul des variations. *J. Math. Pures Appl.* (8), 4 (1921), (67–96).

Work related to Section 7 includes asymptotically optimal polygonal approximation (McClure and Vitale [15]), computational procedures for displaying and analyzing convex sets (Vitale and Tarr [23]), properties of support functions (Vitale [25]), and limit theorems for sequences of random sets (Artstein and Vitale [2] and Vitale [24]).

Sendov [19, 20], has considered the problem of approximation in the Hausdorff metric of sets defined by functions of one real variable and its relationship to the theory of $\epsilon$-entropy.

The article [21] by Ulam advocates a methodology which is similar to the one adopted in this paper. Particularly pertinent are his concepts of "quasi-fixed points" and "$\epsilon$-morphisms."

## ACKNOWLEDGMENTS

We gratefully acknowledge the help and comments received from the following individuals: Dr. Franz Alt, Professor Howard Eves, Dr. Michael Goldberg, and Miss Marion Michaud.

## REFERENCES

1. F. L. ALT, Digital pattern recognition by moments, *J. Assoc. Comput. Mach.* **9** (1962), 240–258.
2. Z. ARTSTEIN AND R. A. VITALE, A strong law of large numbers for random compact sets, *Ann. Prob.* **3** (1975), 879–882.
3. R. P. BOAS, "Entire Functions," Academic Press, New York, 1954.
4. P. J. DAVIS AND H. O. POLLAK, On the analytic continuation of mapping functions, *Trans. Amer. Math. Soc.* **87** (1958), 198–225.
5. P. J. DAVIS, Plane regions determined by complex moments, *J. Approximation Theory* **19** (1977), 148–153.
6. F. DEUTSCH, "Metric Projections," Lecture Notes, 1973, Dept. of Mathematics. Pennsylvania State University.
7. R. O. DUDA AND P. E. HART, "Pattern Classification and Scene Analysis," Chap. 9, Wiley, New York, 1973.
8. H. G. EGGLESTON, "Convexity," Cambridge Univ. Press, London/New York, 1958.
9. V. E. GUILIANO *et al.*, Automatic pattern recognition by a Gestalt method, *Inform. Contr.* **4** (1961), 332–345.
10. G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA, "Inequalities," Cambridge Univ. Press, London/New York, 1967.
11. R. A. JOHNSON, "Advanced Euclidean Geometry," Dover, New York, 1960.
12. M. KAC, Can one hear the shape of a drum? *Amer. Math. Monthly* **73** (1966), 1–23.
13. S. KARLIN AND W. J. STUDDEN, "Tchebycheff Systems: with Applications in Analysis and Statistics," Wiley, New York, 1966.
14. J. L. KELLEY, "General Topology," Van Nostrand, Princeton, 1955.
15. D. E. MCCLURE AND R. A. VITALE, Polygonal approximation of plane convex sets, *J. Math. Anal. Appl.* **51** (1975), 326–358.
16. G. PÓLYA AND G. SZEGÖ, "Isoperimetric Inequalities in Mathematical Physics," Princeton Univ. Press, Princeton, N.J., 1951.
17. G. B. PRICE, On the completeness of a certain metric space with an application to Blaschke's selection theorem, *Bull. Amer. Math. Soc.* **46** (1940), 278–280.
18. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, N.J., 1970.
19. B. SENDOV, Approximation relative to Hausdorff distance, *in* "Approximation Theory" (A. Talbot, Ed.), pp. 101–108, Academic Press, New York, 1970.
20. B. SENDOV, Some questions in the theory of approximation of functions and sets in the Hausdorff metric, *Uspehi Mat. Nauk* **24** (1969), 141–178 (in Russian).
21. S. M. ULAM, On general formulations of simulation and model construction, *in* "Prospects for Simulation and Simulators of Dynamic Systems" (Shapiro and Rogers, Eds.), pp. 3–8, Spartan–Macmillan, New York, 1967.
22. F. A. VALENTINE, "Convex Sets," McGraw-Hill, New York, 1964.
23. R. A. VITALE AND A. TARR, An APL Package for Convex Geometry, Proc. APL-75, Congress, Pisa, 1975, pp. 376–383.

24. R. A. VITALE, Asymptotic area and perimeter of sums of random plane convex sets, Mathematics Research Center Technical Summary Report, to appear.
25. R. A. VITALE, On characterizing support functions of plane convex sets, Div. Appl. Math. Report, Brown University, 1975.
26. M. T. WASAN, "Stochastic Approximation," Cambridge Univ. Press, London/New York, 1969.
27. I. M. YAGLOM AND V. G. BOLTYANSKII, "Convex Figures," Holt, Rinehart & Winston, New York, 1961.